mmEat: Millimeter Wave-Enabled Environment-invariant Eating Behavior Monitoring

Yucheng Xie[a,*], Ruizhe Jiang[a], Xiaonan Guo[a], Yan Wang[b], Jerry Cheng[c], Yingying Chen[d]

[a] *Indiana University-Purdue University Indianapolis, United States*
[b] *Temple University, United States*
[c] *New York Institute of Technology, United States*
[d] *Rutgers University, United States*

Abstract

Dietary habits are closely related to people's health condition. Unhealthy diet can cause obesity, diabetes, heart diseases, as well as increase the risk of cancers. It is necessary to have a monitoring system that helps people keep tracking his/her eating behaviors. Traditional sensor-based and camera-based dietary monitoring systems either require users to wear dedicated devices or may potentially incur privacy concerns. WiFi-based methods, though yielding reasonably robust performance in certain cases, have major limitations. The wireless signals usually carry substantial information that is specific to the environment where eating activities are performed. To overcome these limitations, we propose mmEat, a millimeter wave-enabled environment-invariant eating behavior monitoring system. In particular, we propose an environment impact mitigation method by analyzing mmWave signals in Dopper-Range domain. To differentiate dietary activities with various utensils (i.e., eating with fork, fork and knife, spoon, chopsticks, bare hand) for fine-grained eating behavior monitoring, we construct Spatial-Temporal Heatmap by integrating multiple dimensional measurements. We further utilize an unsupervised learning-based 2D segmentation method and an eating period derivation algorithm to estimate time duration of each eating activity. Our system has the potential to infer the food categories and eating speed. Extensive experiments with over 1000 eating activities show that our system can achieve dietary activity recognition with an average accuracy of 97.5% and a false detection rate of 5%.

Keywords: mmWave sensing, Eating behaviour monitoring

1. Introduction

Dietary is an important activity in people's daily lives since it is closely related to individuals' health conditions. CDC shows that an unhealthy diet can cause obesity, diabetes, heart diseases, as well as increase the risk of over 13 types of cancers [1]. A recent study [2] has shown that unhealthy diet contributes to approximately 678,000 deaths each year in the U.S. Thus, it is necessary to develop a monitoring system that can help individuals keep tracking their dietary behaviors and offer them useful suggestions.

Eating behavior monitoring can provide essential information (e.g., food categories, eating speed) for dietary behavior analysis and provide useful recommendations if poor dietary behaviors are detected. Traditional eating monitoring systems [3, 4] use cameras to take images or videos of users to track their dietary information. However, those vision-based methods may raise potential privacy concerns from collecting images or videos of users. In contrast, some studies [5, 6] propose to use wearable sensors for dietary monitoring. Though sensor-based methods do not raise privacy concerns, users are required to wear one or multiple sensors during eating, which is inconvenient and impractical.

These inconveniences contribute to the emergence of device-free monitoring systems such as WiFi-based methods. Lin et al. propose WiEat [7], which utilizes channel state information extracted from WiFi devices to recognize

*Corresponding author

Email addresses: yx11@iupui.edu (Yucheng Xie), ruizjian@iupui.edu (Ruizhe Jiang), xg6@iupui.edu (Xiaonan Guo), y.wang@temple.edu (Yan Wang), jcheng18@nyit.edu (Jerry Cheng), yingche@scarletmail.rutgers.edu (Yingying Chen)

different dietary activities. However, as WiFi signals are sensitive to surrounding environments and are vulnerable to interference, more stable and stronger signals are desired for eating monitoring tasks. Recent years have witnessed the success of using mmWave signals for posture estimation [8] or activity recognition [9]. This is because mmWave signals have more stable and higher-resolution with shorter wavelengths and stronger directivity. In this paper, we propose to design an eating behavior monitoring system via mmWave techniques which have already been integrated into the new generation WiFi standards (i.e., IEEE 802.11ad).

In order to utilize mmWave signals for eating behavior monitoring, several challenges should be addressed in practice. First, people usually eat in different places (e.g., dining room, living room) every day. Traditional WiFi-based eating monitoring systems that are trained in a specific environment will typically not work well when being applied in a different environment. To solve this problem, in this work, we propose an environmental impact mitigation method by subtracting the static component from every frame in the Dopper-Range domain. Our eating monitoring system is environment-invariant and can be applied to new environments without extra training. Moreover, in the real world, people might perform non-eating activities throughout the day. Hence, we develop a dietary activity detection method to detect eating activities automatically based on the repetitive velocity pattern of eating activity in the time domain. Furthermore, fine-grained eating behavior monitoring requires differentiation among eating activities with various utensils (e.g, eating with a fork or spoon). However, different dietary activities are hard to be distinguished since they all involve hand movements with similar ranges. To address this problem, we construct Spatial-Temporal Heatmap by integrating velocity information from every distance measurement in the Doppler-Range domain and combining them with time information. Besides, we utilize an unsupervised learning-based 2D segmentation algorithm to facilitate accurate dietary activity recognition. We further develop a deep neural network to extract the unique characteristics of every eating activity and classify them based on the utensils used (i.e., fork, fork&knife, spoon, chopsticks, bare hand). In addition, to further derive detailed dietary behavior information, we estimate the eating period of every eating activity and infer the eating duration and speed of meals.

The contribution of our works are summarized as follows:

- As far as we know, mmEat is the first eating behavior monitoring system using COTS mmWave radar sensor.
- Our proposed system constructs unique environment-invariant Spatial-Temporal signal representations that integrate velocity, time duration, and range of movement information.
- Our proposed system has the capability of eliminating environmental impact from static objects and differentiating eating activities from daily activities. Moreover, we develop a fine-tuned deep neural network to facilitate accurate dietary activity recognition.
- Extensive experiments with 6 people over 1000 eating activities show that our system can achieve dietary activity recognition with an average accuracy of 97.5% and a false detection rate of 5%.

2. Related Work

Traditional eating monitoring systems widely use Vision-based methods [3, 4]. Such methods use cameras to take images or videos when users eat meals for further analysis. DietCam [3] exploits photos or videos taken by commercial mobile devices to perform dietary monitoring. Another system developed by O'Loughlin et al. [4] exploits Microsoft SenseCams to capture videos and estimate the dietary energy intake. Such vision-based methods usually raise potential privacy concerns since the camera may capture users' private information such as social relationships and location privacy.

Some existing work tend to use wearable sensors for dietary monitoring to avoid potential privacy concerns in vision-based methods. Amft et al. [5] use a condenser microphone to detect air-conducted vibrations caused by chewing to determine food textures. Zhang et al. [6] propose an accelerometer-based wearable device attached to users' wrists to detect eating activities based on the three-dimensional kinematics movement model. Though sensor-based methods do not have privacy concerns, users are required to wear one or multiple sensors during eating, which is inconvenient and impractical.

Recently, radio frequency (RF) signals have been proposed to address the above limitations. As a prevalent RF sensing modality, WiFi signals have shown initial success in many activity recognition applications. Wang et al.

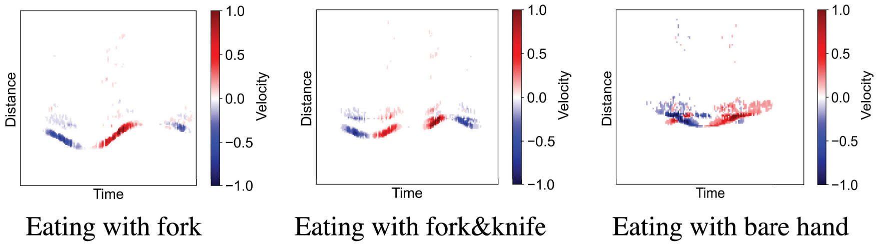

Figure 1: Spatial-Temporal Heatmap of three eating activities.

develop E-eyes [10], which exploits WiFi signals to provide device-free human activity identification. Lin et al. develop WiEat [7] that can achieve high accuracy in device-free dietary monitoring using commercial WiFi devices. However, WiFi-based methods are sensitive to environmental changes. Millimeter wave (mmWave) has been proven more robust than WiFi due to its high bandwidth and native beam-forming technology. Existing mmWave-based systems like [8] and [9] often focus on posture estimation or activity recognition. None of them show that their system can distinguish eating activities with minute differences in hand or finger movements and provide fine-grained analysis of eating activities. This paper develops a system leveraging mmWave signals from commodity mmWave devices to provide fine-grained dietary monitoring.

3. System and Methodology

3.1. Preliminaries

The intuition behind monitoring eating activity using mmWave is that eating activities with different utensils have minute but different action components. For example, "eating with fork and knife" has a cutting action while "eating with fork" does not. Such action components generate different reflections of mmWave signals that can be utilized for eating activity monitoring. To demonstrate the feasibility of dietary activity recognition, we conducted experiments by asking one participant to perform 3 dietary activities with different utensils (e.g., eating with fork, fork&knife, bare hand) in an office. Specifically, a mmWave device (i.e., AWR1642) with a sampling rate of 100 frames/sec is placed at one end of a table. The participant sits in front of the table with $1m$ away from the device while performing these activities. As shown in Figure 1, the three Spatial-Temporal Heatmap of velocity, distance, and time duration have significantly different patterns for the three eating activities.

3.2. System Overview

The goal of mmFit is to provide environment-invariant fine-grained eating behavior monitoring by leveraging a single commercial mmWave device. Toward this end, we develop a low-cost mmWave-based eating behavior monitoring system, mmEat. The system takes as input the mmWave signals reflected from the human body. The system first performs signal processing to derive the velocity, distance information of the user's activity from the received mmWave signals. Then, it eliminates the impact from environment by subtracting signals reflected off static objects. Next, we construct Spatial-Temporal Heatmap to aggregate the instantaneous velocity from every distance measurement in the Doppler-Range domain and combine them with time information. Such integrated multidimensional signal representation can facilitate fine-grained activity recognition. We propose a dietary activity detection method based on the repetitive eating activity patterns in the time domain to detect dietary activities based on the Spatial-Temporal signal representation. To further differentiate eating activities, we apply DBSCAN [11] to cluster and segment each activity, and develop a deep neural network to identify them. The last component of our proposed system is eating period monitoring which estimates the eating period of each eating activity. Such information is useful to assist various health-related problems, such as diabetes, heart diseases, etc. The overview of mmEat is shown in Figure 2.

3.3. Spatial-Temporal Signal Representation

Signal Preprocessing. We first perform range-FFT and Doppler-FFT on the received mmWave signals to derive the distance and velocity information of user's activity respectively. Then, we derive the Doppler-Range Heatmap based on the instant velocity and distance measurements. As shown in Figure 3, the heatmap indicates the strength of frequency responses of the reflected signals via the color. However, since static objects (e.g., furniture and walls) in the

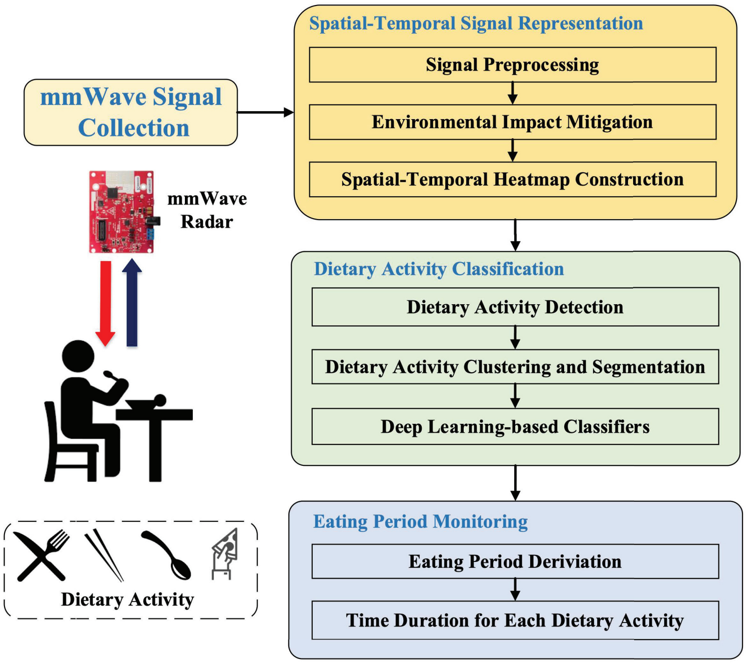

Figure 2: System overview of mmEat.

environments can also reflect mmWave signals, it is still hard to extract signals from the human in the Doppler-Range Heatmap.

Environmental Impact Mitigation. To eliminate the environmental impact mentioned above, we propose an environmental impact mitigation method by filtering out of non-moving objects in Doppler-Range domain. We note that the frequency responses of the mmWave signals reflected from static objects in the environment (e.g., walls and furniture) do not change over time. Therefore, we can eliminate the impact caused by static objects by subtracting the time-invariant frequency response from the Doppler-Range Heatmap. In particular, we collect mmWave signals in a static environment for a short period (e.g., 3 min) and derive the Doppler-Range Heatmap to estimate the time-invariant frequency response.

Spatial-Temporal Heatmap Construction. Although the denoised Doppler-Range Heatmap can capture the instant velocities at different distances, it is not enough to describe the process of the dietary activities. We propose a more comprehensive signal representation by constructing the Spatial-Temporal Heatmap that contains the temporal information of eating activities (e.g., time duration of each activity and variation of velocity with time). Specifically, we accumulate the velocity measurements of each distance in every Doppler-Range Heatmap frame and then present their dynamics in the time domain as follows:

$$V_{q,t} = \sum_{p=1}^{D}(f_{p,q,t}) \times v_{p,t}, p \in [1, D], q \in [1, R], \tag{1}$$

where $f_{p,q,t}$ is the strength of a frequency response in the Doppler-Range Heatmap, p is the doppler index, q is the range index, and t is the frame index. $v_{p,t}$ is the velocity corresponding to a Doppler index p at frame t. Then we normalize the derived $V_{q,t}$ to $[-1, 1]$ and map the original 2-dimensional Doppler-Range data to a more comprehensive 3-dimensional Spacial-Temporal Heatmap, which presents the process of the eating activities as shown in Figure 4.

3.4. Dietary Activity Classification

Dietary Activity Detection. After constructing Spatial-Temporal signal representations from mmWave signals, we perform the dietary activity detection to determine whether the mmWave signals contain dietary activities or not. We find that dietary activities usually have repetitive patterns in the Spatial-Temporal domain while non-dietary activities do not. The reason is that dietary activities consist of repetitive hand and arm movements that bring food to

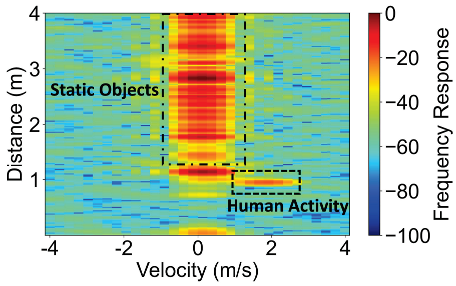

Figure 3: Doppler-Range Heatmap when one user is eating using a fork.

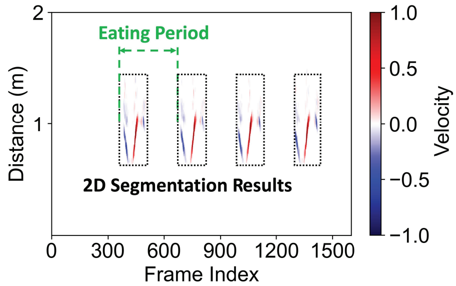

Figure 4: Spatial-Temporal Heatmap of four eating activities using a fork.

the mouth from the table. Based on the observation, we propose to detect dietary activities by searching the repetitive patterns in the Spatial-Temporal Heatmap using a sliding window. Particularly, we accumulate the velocities from all the distances at a particular time in a frame to transfer the heatmap to one-dimensional data. We use an autocorrelation-based method to determine whether the mmWave signals contain a repetitive pattern or not. We empirically determine that a dietary activity is detected when the number of peaks in the autocorrelation results is more than 5.

Dietary Activity Clustering and 2D Segmentation. Once a dietary activity is detected, the system performs the dietary activity segmentation to focus on the signals related to dietary activities. The basic idea is to determine each dietary activity's time duration and range of movement in the Spatial-Temporal Heatmap. We first remove the points with low absolute velocity from the heatmap based on an empirical threshold. Then, we utilize an unsupervised learning-based clustering method (i.e., DBSCAN) to separate the points into different clusters. We design a dynamic algorithm to determine the 2D window size of each activity based on its time duration and range of movement. Particularly, for each cluster, we determine the window size based on the differences between the coordinates of the edge points in the Spatial-Temporal plane. The box in Figure 4 illustrates the 2d segmentation results of our algorithm. In addition, we scale up the size of the window by an empirical constant (i.e., 1.2) to ensure that it contains all the signals related to dietary activities.

Deep Learning-based Classifier. We choose to use neural network-based method for final classification since it has shown robust performance in image classification tasks [12]. The segments derived by the proposed segmentation method are first resized to images with size of 224×224. Our convolutional neural network contains 9 layers. 3 convolutional layers are exploited for up-sampling, 3 Max Pooling layers with each follows a convolutional layer are used for down-sampling. After the process of 3 rounds of up-sampling and down-sampling, a 64-dimensional feature map is obtained and a flatten layer is followed to reduce the feature map into a one-dimension array. Two dense layers at the end of the network will classify arrays into 5 categories, each category is mapped to a specific dietary activity.

3.5. Eating Period Monitoring

Researchers [13] have demonstrated that the speed of eating is an important factor for weight control. People eating quickly have a significantly higher possibility of obesity. The basic idea of eating period monitoring is to derive the accurate time duration of each eating activity and infer detailed eating information (e.g., eating period of a meal, eating speed). Given that objective, we propose an eating period derivation method. We infer the time duration of each eating activity based on calculating the interval with neighboring activities. Specifically, as shown in Figure 4, we determine the beginning of each eating activity by searching the time stamp of the left edge from the 2D segmentation box. We then estimate the eating period of each eating activity based on the differences between consecutive time stamps. By estimating the time duration of each eating activity, we could further infer the accumulated eating period using specific utensils during a meal, which could be used to estimate other high-level information such as the calorie intake and nutrition balance. In addition, the number of eating activities during a meal and average eating period could also be used to detect poor dietary behaviors of users, such as overeating and eating too quickly.

4. Performance Evaluation

4.1. Experimental Setup:

Devices: In our experiments, we use a single TI AWR1642 commercial mmWave radar equipped with a 2×4 antenna array. The radar operates at a frequency band between 77GHz and 81GHz with a sampling rate fixed at 100

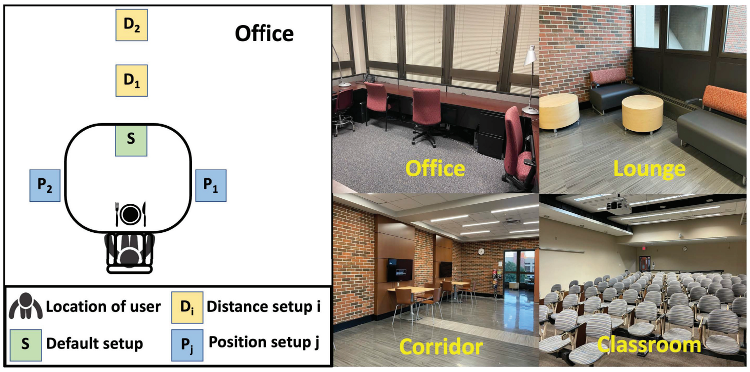

Figure 5: Experiment Setup.

frames per second. All devices are attached to a DELL G3 laptop for deep learning model inference.

Data Collection: We conduct the experiments by recruiting 6 volunteers (age from 22 to 40). The profiles are collected at an office with a size of $5 \times 3\ m^2$. A total of 5 typical eating activities are performed by the volunteers. Over 1000 eating activity data are collected and the ground truths are measured and verified by camera-based method during the experiments. As is shown in Figure 5, we totally test three different positions and three distances ($1m$, $1.5m$ and $2m$) to evaluate impact of device positions and distances. For the evaluation of environment impact, we collect data under three different environments: A). a lounge with a size of $4 \times 4\ m^2$; B). a corridor with a size of $5 \times 9\ m^2$; C). a classroom with a size of $9 \times 15\ m^2$.

Evaluation Metrics: We define four different evaluation metrics: *Dietary Activity Recognition Accuracy* is the percentage of predicted dietary activities that are correctly recognized among all activities; *False Detection Rate (FDR)* is defined as the ratio between the number of incorrectly classified activities and the total number of activities. *Confusion Matrix* visualizes the percentage of a specific activity being identified among all the activities. *Estimated Error* defines the difference between the estimated eating duration and actual eating duration for a single dietary activity.

4.2. Performance of Dietary Activity Classification

In this section, we first compare the overall performance of the proposed CNN-based classification method with traditional classifiers. Figure 6 demonstrates the overall recognition accuracy and FDR of five classifier. Our CNN-based method outperforms all four traditional methods and achieves 96.78% in recognition accuracy and 3.3% in FDR. We then show the dietary activity classification for five activities. As shown in Figure 7, the recognition accuracy for all activities are higher than 90%. The accuracy of using bare hand is a little lower than other activities, because the body movement of using bare hand is similar to that of using spoon, which may cause some confusion to the classifier. The result confirms that our CNN-based classifier can achieve robust performance in dietary activity classification.

4.3. Impact of Different Environments

We then evaluate the impact of different environments on system performance. In particular, we collect data from three different environments mentioned in section 4.1. We use data from one environment as the training set and data from the other two environments as the testing set and try different training-testing pairs. As demonstrated in Figure 8, all of the training-testing pairs achieve classification accuracy over 88% and with FDR below 9% even the training set and testing set are collected from different places. This result proves that our system is able to offer domain-invariant performance under different environments.

4.4. Impact of Different Device Position

Different positions of the device may affect the accuracy of dietary activity classification. We study the impact of device position on our CNN-based classifier. We evaluate three positions demonstrated in Figure 5. As shown in Figure 9, at all three positions, our system maintains an FDR lower than 4%. The accuracy of position P1 and position P2 are slightly lower than that of default position S. This is because when the device is located at a position not facing the user, user's arms are parallel to the device, causing weaker Doppler effects and vaguer Spatial-Temporal Heatmaps. But our system still maintains an accuracy over 94%. The result proves that our system can still maintain

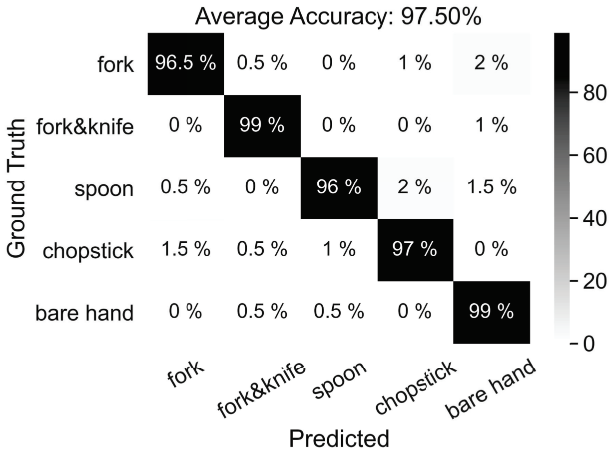

Figure 6: Confusion matrix of dietary activity classification using CNN-based classifier.

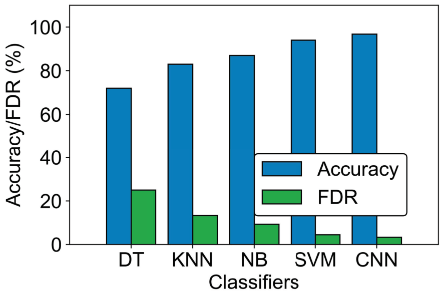

Figure 7: Performance Comparison among four traditional machine learning and CNN-based classifier.

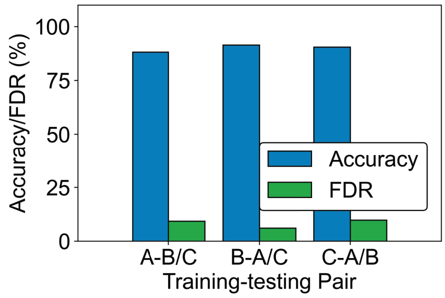

Figure 8: Impact of environment.

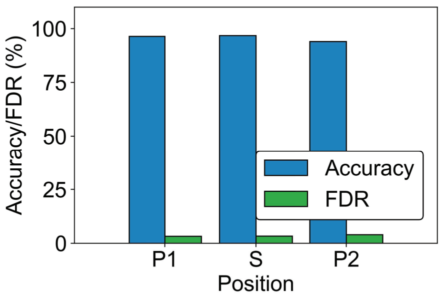

Figure 9: Impact of different positions of mmWave device in office.

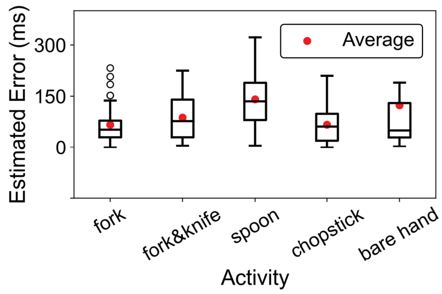

Figure 10: Comparison of eating period estimated error.

a good performance in dietary activity classification even the device is situated at different positions. We also test the system at three distances (i.e., 1m, 1.5m, 2m) at position S and find that the performance is not affected.

4.5. Performance of Eating Period Monitoring

We next evaluate the performance of eating period monitoring for different food intake activities. In our experiments, each of the 5 activities is performed 160 times and we collected 800 eating activities in total. As shown in Figure 10, the average estimated error (indicated by red points) of using fork, fork&knife, spoon, chopsticks and bare hand are $67ms$, $88ms$, $141ms$, $67ms$ and $124ms$, respectively, which are all within $150ms$. Additionally, the estimated error for all the collected activities are all smaller than $400ms$. The results demonstrate that our proposed system can precisely estimate eating period and maintain a low estimated error for different activities. Furthermore, by calculating the average time duration of each eating activity, we can estimate users' eating speed and infer high-level information such as calories intake or analysis of nutritional balance. The detailed dietary information can be further used to assist the healing of various health problems caused by bad eating habits.

5. Conclusion

In this paper, we explore the feasibility of using mmWave signals for fine-grained dietary behavior monitoring. We show that the proposed CNN-based eating behavior monitoring system is environment-invariant and can be applied to new environments without extra training efforts. We also demonstrate the potential of the proposed system to provide users with comprehensive understanding of their eating behaviors and help them get rid of unhealthy dietary habits. Extensive experimental results show that our system can achieve dietary activity recognition with over 97.5% average accuracy and less than 5% FDR.

6. Acknowledgments

This work was partially supported by National Science Foundation Grants CNS-2120396, CCF-2028876, ECCS-2033433, CCF-2000480, CNS-2120276, CCF-2028858, CNS-2120371, CNS-1717356, CCF-2028894, CCF-2028873, CCF-1909963, CNS-2120350.

References

[1] The harmful effects of poor nutrition, https://www.cdc.gov/chronicdisease/resources/publications/factsheets/nutrition.htm.

[2] Why good nutrition is important, https://www.cspinet.org/eating-healthy/why-good-nutrition-important.

[3] F. Kong, J. Tan, Dietcam: Automatic dietary assessment with mobile camera phones, Pervasive and Mobile Computing 8 (2012) 147–163. doi:10.1016/j.pmcj.2011.07.003.

[4] Using a wearable camera to increase the accuracy of dietary analysis (2013). doi:10.1016/j.amepre.2012.11.007.

[5] O. Amft, A wearable earpad sensor for chewing monitoring, in: SENSORS, 2010 IEEE, 2010, pp. 222–227. doi:10.1109/ICSENS.2010.5690449.

[6] S. Zhang, M. Jr, W. Xiao, C. K. Tham, Detection of activities by wireless sensors for daily life surveillance: Eating and drinking, Sensors (Basel, Switzerland) 9 (2009) 1499–517. doi:10.3390/s90301499.

[7] Z. Lin, Y. Xie, X. Guo, Y. Ren, Y. Chen, C. Wang, Wieat: Fine-grained device-free eating monitoring leveraging wi-fi signals, in: 2020 29th International Conference on Computer Communications and Networks (ICCCN), 2020, pp. 1–9. doi:10.1109/ICCCN49398.2020.9209628.

[8] M. Raja, Z. Vali, S. Palipana, D. G. Michelson, S. Sigg, 3d head motion detection using millimeter-wave doppler radar, IEEE Access 8 (2020) 32321–32331. doi:10.1109/ACCESS.2020.2973957.

[9] Y. Ren, J. Lu, A. Beletchi, Y. Huang, I. Karmanov, D. Fontijne, C. Patel, H. Xu, Hand gesture recognition using 802.11ad mmwave sensor in the mobile device, in: 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), 2021, pp. 1–6. doi:10.1109/WCNCW49093.2021.9419978.

[10] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, H. Liu, E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures, in: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, MobiCom '14, Association for Computing Machinery, New York, NY, USA, 2014, p. 617–628. doi:10.1145/2639108.2639143.
URL https://doi.org/10.1145/2639108.2639143

[11] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, X. Xu, Dbscan revisited, revisited: why and how you should (still) use dbscan, ACM Transactions on Database Systems (TODS) 42 (3) (2017) 1–21.

[12] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition (2015). arXiv:1409.1556.

[13] T. Ohkuma, Y. Hirakawa, U. Nakamura, Y. Kiyohara, T. Kitazono, T. Ninomiya, Association between eating rate and obesity: a systematic review and meta-analysis, International journal of obesity 39 (11) (2015) 1589–1596.